DOCUMENT RESUME

ED 274 677 TM 860 536

AUTHOR Furtwengler, Carol; And Others
TITLE Multiple Data Sources in Teacher Evaluation.
PUB DATE Apr 86
NOTE 11p.; Paper presented at The National Council on Measurement in Education (San Francisco, CA, April 17-19, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Career Ladders; Classroom Observation Techniques; *Data Collection; Elementary Secondary Education; *Evaluation Criteria; Evaluation Methods; Merit Pay; *Peer Evaluation; Portfolios (Background Materials); Questionnaires; Scoring; State Programs; *Teacher Evaluation; Teacher Promotion
IDENTIFIERS Tennessee; *Tennessee Career Ladder Program

ABSTRACT

The Tennessee Career Ladder Program is a statewide evaluation system designed to pay teachers for teaching well. The system is based on a set of objectives called Domains of Competence, which are subdivided into Indicators. Early in the design of the system, it was decided to use multiple data sources, including a variety of instruments, to assess the Domains and Indicators. This paper reports the 1984-85 evaluation of 3,000 experienced classroom teachers. Grade levels K-12 and general education subject areas were included. Each teacher was evaluated by a team of three trained peer teachers from other school systems. The 105 full-time evaluators received more than 60 hours of training in administering the seven instruments: classroom observation, portfolio analysis, principal questionnaire, student questionnaires, peer questionnaires, a test, and evaluator judgment. Analysis includes descriptions of the evaluation system and of changes made in the system for use in 1985-86 and the reasons for them. The Domains and Indicators used in the 1984-85 evaluation and their data sources are included in the appendix. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED274677

Multiple Data Sources In Teacher Evaluation

Carol Furtwengler
George Malo
Joyce McLarty
Sara Strouss

Department of Education
State of Tennessee

Presented to
The National Council on Measurement in Education
San Francisco, CA
April, 1986

Running Head: Multiple Data Sources

TM 860 536

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. McLarty

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

Multiple Data Sources in Teacher Evaluation

This paper reports the 1984-85 evaluation of 3,000 experienced classroom teachers. Grade levels K-12 and general education subject areas were included. Each teacher was evaluated by a team of three trained peer teachers from other school systems. The 105 full-time evaluators received more than 60 hours of training in administering the seven instruments: classroom observation, portfolio analysis, principal questionnaire, student questionnaires, peer questionnaires, a test, and evaluator judgment. Analysis includes descriptions of the evaluation system and of changes made in the system for use in 1985-86 and the reasons for them.

MULTIPLE DATA SOURCES IN TEACHER EVALUATION

The Tennessee Career Ladder program is a statewide system designed to pay teachers more for teaching well. The evaluation upon which the determination of meritorious performance is based focuses on a set of objectives called Domains of Competence. The Domains are subdivided into Indicators of Competence, and these are still further described by measurement statements.

Early in the design of the Career Ladder evaluation system, it was decided to use multiple data sources to assess these domains and indicators. This was because each type of assessment (tests, interviews, observations, questionnaires) has certain flaws, and because the different objectives did not lend themselves to a uniform method of assessment. Classroom management, for example, might better be assessed by a classroom observation than an interview. Planning, however, might show more clearly in a portfolio analysis than in a questionnaire. The Domains and Indicators used in the 1984-85 teacher evaluation system and their data sources are included in the Appendix.

To arrive at the fairest and most accurate evaluation for each teacher, each objective was assessed using at least three data sources. This videotape explains how these data sources are administered.

**********insert videotape here**********

What you have just seen explains the Career Ladder system as it is being used in 1985-86. Several improvements have been made since 1984-85. Each of these improvements is based on things learned in the first year of implementation of the Career Ladder evaluation system. These changes, and the reasons for them, constitute a concise description of the most important findings from the first year. In keeping with the general approach used in the Career Ladder program, both quantitative and qualitative information were used in making decisions about appropriate modification. Changes are summarized in Table 1.

Classroom Observations

Two changes were made in classroom observations: the number of observations was doubled (from 3 to 6) and the proportion of unannounced visits was increased from 1 out of 3 to 2 out of 3. It was found that teacher activities varied from one visit, subject area, type of lesson and group of students to another. Added observation time was, therefore, statistically appropriate. In addition, both candidates and evaluators favored having the evaluator spend more time in the classroom.

Table 1

BEFORE and AFTER

CAREER LADDER TEACHER EVALUATION SYSTEM

Improvement for 1985-86

A review and revision of the Career Ladder Teacher Evaluation System for 1984-85 has resulted in changes to make the system even better during 1985-86. A comparison of the system for 1984-85 with 1985-86 is shown below.

LAST YEAR	THIS YEAR
Classroom Observations -three (per 3 visits) -two announced -one unannounced	Classroom Observations -six (per 3 visits) -two announced -four unannounced
Questionnaires -Student K-2 9 items 3-6 39 items 7-12 44 items -Superordinate* 45 items no explanation -Peer 20 items	Questionnaires -Student K-2 9 items 3-6 34 items 7-12 39 items -Principal* 20 items explanation -Peer 16 items
Portfolio -Includes four areas -Rated at school by an evaluator -Verification required	Leadership Summary -Includes one area -Submitted "B" visit -Rated centrally -Only candidate signature required
Interview (eliminated) -Informal sharing	Dialogue Sessions** -Planning -Teaching strategies -Evaluation
Evaluator Consensus -5 Domain Scores	Evaluator Consensus -4 Domain Scores -Not in Leadership Domain***

LAST YEAR	THIS YEAR
Student Performance -Not considered	Student Performance -Field test only
Instrumentation -Confidential	Instrumentation -Open (included in manual)
Orientation Manual	Orientation Manual -Revised, expanded
	Principal's Handbook -Improve Communication -Explain Questionnaire -Explain Role of Principal
	Evaluator's Handbook -Improve Communication
Career Ladder Test	Career Ladder Test -No changes
Scoring -Ten cut-off scores	Scoring -Five minimum scores -Two cut-off scores -Standards equated to 1984-85 system

*The 1984-85 Questionnaire used a frequency scale which asked how often a teacher did something. The 1985-86 Questionnaire uses a quality scale which asks how well a teacher does something.

**Combination of interview with analysis of materials formerly in portfolio

***No consensus decision score will be assigned by the evaluators for the Leadership Domain.

The State Certification Commission and State Board of Education have the right to improve the certification system and make appropriate changes in evaluation standards, criteria, and procedures as needed to maintain a high quality equitable and effective certification system.

Both candidates and evaluators also recommended having two of the three visits unannounced. Some candidates disliked the waiting and worrying they experienced before an announced visit. Both groups felt that a good teacher is good at any time.

Questionnaires

All of the questionnaires had high internal consistency reliability. It was therefore possible to streamline all but the K-12 student questionnaire (which had only 9 items) by eliminating some redundant items.

The superordinate (principal) questionnaire was substantially revised twice. The first revision, as indicated in the table, moved the instrument from a frequency scale (supplying data) to a quality scale (supplying a judgment). Analysis of the 1984-85 data showed severe negative skews for all principal questionnaire items, and it was hoped that this change would improve both the items' discrimination and the respondents' understanding of the instrument.

After considerable discussion, the scale was further modified to allow a direct conversion of the principal's judgment to the final reporting scale (200-800). This step was undertaken when it became apparent that principals were unwilling to assign teachers anything less than the highest possible score for fear that the effect of conversion onto the reporting scale would be other than the one they intended. The revised scale allows them to "vote" on the final status of the candidate.

Portfolio and Interview

The structured candidate interview developed for 1984-85 was dropped partway through the year after the confidentiality of the secure rating scales was breached. Portfolio analysis was also affected to some extent by the release of secure rating scales. In addition, teachers were found to spend substantial time in portfolio preparation and this was neither an intended nor a desired outcome of this aspect of the assessment.

Elements of the portfolio analysis and interview were, therefore, combined to create a series of three Dialogues covering planning, teaching strategies, and evaluation. Each evaluator conducts one hour-long dialogue with the candidate, consisting of 30 minutes of structured questions and 30 minutes during which the candidate may share whatever he or she wishes. The structured items are scored; the sharing time is considered during consensus judgment.

One element of the old portfolio continues to be submitted for analysis: the leadership summary. This resume-like document is scored centrally to enhance scoring accuracy and efficiency.

Evaluator Consensus

Evaluator consensus was eliminated for the leadership domain since evaluators would no longer be scoring the portfolio in which this information was included.

Student Performance

Not considered in 1984-85, the level of student performance was included for field test purposes only in 1985-86. The approach taken was to ask the candidate to provide evidence of student academic growth and of improvement of student attitudes toward learning. Relevant questions were also included in the principal and student questionnaires.

Although there was general agreement that student performance, in its broadest sense, is the intended outcome of instruction, the method of assessing it and the degree to which teachers should be held accountable for it are still topics of much concern and discussion. The field-test data are intended to help map directions for future years.

Instrumentation

In 1984-85, most instrumentation was confidential. Teachers were informed of the Domains and Indicators to be assessed, but not permitted to inspect the instruments. This led to some mistrust on the part of the teachers and, indirectly, to the release of secure rating scales. This, in turn, forced dropping of the interview and a reduction of weight on the portfolio.

Currently only the written test items remain confidential. As with most changes, this solves some problems and creates others. To the degree that quantitative instruments and detailed scoring criteria are available to the candidates, the criteria tend to modify the candidate's teaching behavior. If this causes a candidate to adopt improved instructional methods, it is a positive outcome of the evaluation process. If it results in a candidate restricting a successful teaching method in order to mimic a single model of good teaching, however, it is counter- productive.

The issues surrounding: (1) the teacher's right to know the basis of the evaluation, (2) the difficulty of maintaining security of any kind in a statewide system, (3) the degree to which evaluations are to be objective (quantitative) versus subjective (qualitative), and (4) the degree to which the evaluation can or should dictate teaching style, are likely to become of paramount importance this year.

Orientation Manual

The original 1984-85 orientation manual has been expanded and revised to improve readability. Additional communications in the form of a Principal's Handbook, Evaluator's Handbook, and frequent letters to and publications for candidates and others have helped to improve the efficient dissemination of correct information.

Career Ladder Test

No changes were made in the Career Ladder Test, which covers Reading, Writing, and Professional Skills. Candidates were, and are, allowed to retake the Professional Skills portion whenever the test is offered and to submit their best scores as part of the evaluation. Once passed, the Reading and Writing sections need not be retaken.

Scoring

Scoring was changed from using separate cutoff scores in each of the five domains, to minimums to be attained in each domain plus cutoff scores on a single weighted total score. Teachers and evaluators felt that some domains should be more heavily weighted than others. In addition, data analysis revealed that, while the instruments had high internal consistancy reliability, scores from different instruments were not closely related to one another. Factor analysis partially supported the domain structure. The generation of a total score was, therefore, seen as appropriate for both conceptual and statistical reasons.

In order to assure fairness to both 1984-85 and 1985-86 candidates, standards for the two years are to be equated. Because of the various improvements in the instruments, a common items methodology is being employed.

Multiple Data Sources

The continued use of a multiple data source system was supported by correlational and factor analytic studies. Because it appears that the success of a teacher in an evaluation process may depend in part on the instrumentation selected, it is fairest to teachers to use a variety of instruments. This avoids providing some teachers with consistent advantages and others with consistent disadvantages based on individual ability to demonstrate merit on a particular type of instrumentation.

The Career Ladder program is now in its second full year of implementation. In addition to the original teacher system, evaluation systems have been developed for principals and assistant principals, instructional supervisors, special education teachers who do not teach self-contained classes, vocational education teachers, library and media specialists, and counselors. Many other systems are in various stages of competency development, instrument development, pilot testing, field testing, and analysis.

The most successful aspects of the Tennessee Career Ladder program to date have been the use of highly trained peer evaluators from school systems other than the candidate's, the objectives-based format, the blend of data and evaluator judgment, and the use of multiple sources of information. These are choices which are especially recommended to others developing Career Ladder programs.

Appendix

Data Sources for

COMPETENCIES AND INDICATORS

	Classroom Observation	Portfolio	Applicant Interview	Principal Interview	Peer Questionnaire	Student Questionnaire	Written Test
I. Prepares for instruction effectively.							
A. Establishes appropriate instructional goals and related objectives consistent with the curriculum.		X	X	X			
B. Prepares instructional plans and materials incorporating principles of effective instruction.	X	X	X	X			
C. Creates, selects or modifies instructional plans and materials to accommodate learner instructional levels.	X						
II. Uses teaching strategies and procedures appropriate to the content, objectives, and learners.							
A. Provides a clear description of the learning task and its content.	X	X				X	
B. Monitors learner understanding and reteaches as necessary.	X	X	X			X	
C. Provides learners appropriate practice and review.	X	X	X			X	
D. Establishes and maintains learner involvement in the learning task.	X		X			X	
III. Uses evaluation to improve instruction.							
A. Uses information about learner performance to improve the instructional process.		X	X	X			
B. Reports learner status and progress to learners and their parents.	X	X	X	X		X	
IV. Manages classroom activities effectively.							
A. Establishes and maintains appropriate learner behavior.	X	X		X		X	
B. Establishes and maintains a classroom climate conducive to learning.	X	X	X	X		X	
C. Makes effective use of classroom resources (e.g. personnel, time, materials, facilities)	X		X	X		X	
V. Establishes and maintains a professional leadership role.							
A. Improves professional skills and knowledge.		X	X	X			
B. Takes a leadership role in improving education.		X	X	X	X		
C. (Screening only) Performs professional responsibilities efficiently.					X		
VI. (Screening only) Communicates effectively.							
A. Writes clearly and correctly.	X					X	X
B. Communicates oral information effectively.	X					X	
C. Reads professionally relevant literature/materials with comprehension.							

BEST COPY AVAILABLE